What is claimed is:

1. An apparatus for reproducing a music sound and a voice sound, comprising:

a first storing section that stores a music data file containing a music part and a voice part, the music part containing a sequence of music generation events effective to instruct generation of the music sound, the voice part containing voice reproduction sequence data composed of a combination of voice reproduction event data and duration data, the voice reproduction event data instructing reproduction of a sequence of voice events, the duration data specifying a timing of effecting a voice event in terms of a duration time measured from another voice event preceding to the voice event;

a control section that reads out the music data file from the first storing section; and

a sound generator section that operates based on the music part contained in the read music data file for generating the music sound representative of the sequence of the music events, and that operates based on the voice part contained in the read music data file for generating the voice sound representative of the sequence of the vice events, thereby mixing and outputting the music sound and the voice sound.

2. The apparatus according to claim 1, wherein the voice

reproduction sequence data contains formant control information for generating formants of the voice sound, and the voice reproduction event data contained in the voice part of the read music data file instructs reproduction of the formant control information, so that the sound generator section operates based on the formant control information which is contained in the voice reproduction sequence data and which is specified by the voice reproduction event data for generating the voice sound.

3. The apparatus according to claim 1, further comprising a second storing section that stores first dictionary data which records correspondence between text information representing words to be pronounced as the voice sound and phoneme information representing phonemes of the words, and correspondence between prosodic symbols representing vocal expressions applied to pronunciation of the words and prosodic control information for controlling the vocal expressions, and a third storing section that stores second dictionary data which records correspondence between a combination of the phoneme information and associated prosodic control information representing the voice sound to be reproduced, and formant control information used for generating formants of the voice sound, wherein the control section reads out the music data file having the voice part containing the voice reproduction event data of a text description type which instructs reproduction of the voice

sound represented by the text information and associated prosodic symbols, then the control section refers to the first dictionary data stored in the second storing section for acquiring therefrom the phoneme information and associated prosodic control information corresponding to the text information and associated prosodic symbols, and further refers to the second dictionary data stored in the third storing section for reading out therefrom the formant control information corresponding to the acquired phoneme information and associated prosodic control information, so that the sound generator section operates based on the read formant control information for generating the voice sound.

4. The apparatus according to claim 1, further comprising a second storing section that stores dictionary data which records correspondence between a combination of phoneme information and associated prosodic control information, and formant control information, the phoneme information representing phonemes of the voice sound to be reproduced, the associated prosodic control information being capable of controlling vocal expressions of the phonemes, the formant control information being capable of generating formants of the voice sound, wherein the control section operates when the voice reproduction event data contained in the voice part of the read music data file instructs reproduction of information of a phoneme description type containing the phoneme information and associated prosodic control

information corresponding to the voice sound to be reproduced, for referring to the dictionary data stored in the second storing section to acquire therefrom the formant control information corresponding to the phoneme information and associated prosodic control information which are specified by the voice reproduction event data, so that the sound generator section operates based on the acquired formant control information for generating the voice sound.

5. The apparatus according to claim 1, wherein the first storing section stores the music data file containing the voice part of a first format type, the sound generator section is operable based on the voice part of a second format type for generating the voice sound, and the control section detects a format type of the voice part read from the first storing section and operates if the detected first formant type of the voice part is not compatible with the second format type for converting the read voice part from the first format type to the second format type, thereby enabling the sound generator section.

6. The apparatus according to claim 5, further comprising a second storing section that stores dictionary data required for conversion of the format type of the voice part of the music data file, so that the control section refers to the dictionary data stored in the second storing section for effecting the conversion of the format type of the voice part.

7. The apparatus according to claim 1, wherein the voice part of the music data file contains data specifying a kind of language of the voice part.

8. The apparatus according to claim 1, wherein the sound generator section operates based on the voice part of the music data file for generating the voice sound representative of a human voice.

9. A memory medium for storing voice reproduction sequence data designed for causing a sound generator device to reproduce a human voice, wherein

the voice reproduction sequence data has a chunk structure composed of a content information chunk containing information for managing the voice reproduction sequence data and at least one track chunk containing voice sequence data, and wherein

the voice sequence data comprises a sequence of pairs of voice reproduction event data and duration data, the voice reproduction event data instructing a voice reproduction event of the human voice, the duration data specifying a timing of executing the voice reproduction event in terms of a duration time measured from a preceding voice reproduction event.

10. The memory medium according to claim 9, wherein the

voice reproduction event data is one of a text description type, a phoneme description type and a formant frame description type, the text description type of the voice reproduction event data containing text information specifying words to be pronounced by the sound generator device as the human voice and associated prosodic symbols specifying vocal expression applied to pronunciation of the words, the phoneme description type of the voice reproduction event data containing phoneme information specifying phonemes of the human voice to be reproduced by the sound generator device and associated prosodic control information controlling vocal expressions of the phonemes, the formant frame description type of the voice reproduction event data containing formant control information specifying formants of the human voice at respective time frames.

11. A memory medium for storing sequence data for causing a sound generator device to reproduce a music sound and a human voice, wherein the sequence data has a data structure composed of music sequence data and voice reproduction sequence data,

the music sequence data comprising a sequence of pairs of music generation event data and duration data, the music generation event data instructing a music generation event of the music sound, and the duration data specifying a timing of executing the music generation event in terms of a duration time measured from a preceding music generation event, and

the voice reproduction sequence data comprising a sequence of pairs of voice reproduction event data and duration data, the voice reproduction event data instructing a voice reproduction event of the human voice, and the duration data specifying a timing of executing the voice reproduction event in terms of a duration time measured from a preceding voice reproduction event, whereby the music sequence data and the voice reproduction sequence data are concurrently processed by the sound generator device so as to reproduce the music sound and the human voice along a common time axis.

12. The memory medium according to claim 11, wherein the sequence data has a chunk structure such that the music sequence data and the voice reproduction sequence data are arranged at different chunks.

13. The memory medium according to claim 11, wherein the voice reproduction event data is one of a text description type, a phoneme description type and a formant frame description type, the text description type of the voice reproduction event data containing text information specifying words to be pronounced by the sound generator device as the human voice and associated prosodic symbols specifying vocal expression applied to pronunciation of the words, the phoneme description type of the voice reproduction

event data containing phoneme information specifying phonemes of the human voice to be reproduced by the sound generator device and associated prosodic control information controlling vocal expressions of the phonemes, the formant frame description type of the voice reproduction event data containing formant control information specifying formants of the human voice at respective time frames.

14. A server apparatus comprises a storing section and a transmitting section, wherein

the storing section stores a music data file containing a music part and a voice part, the music part containing a sequence of music generation events effective to instruct generation of the music sound, the voice part containing voice reproduction sequence data composed of a combination of voice reproduction event data and duration data, the voice reproduction event data instructing reproduction of a sequence of voice events, the duration data specifying a timing of effecting a voice event in terms of a duration time measured from another voice event preceding to the voice event, and

the transmitting section responds to a request from a client terminal apparatus for distributing the stored music data file to the client terminal apparatus.

15. The server apparatus according to claim 14, wherein the voice reproduction event data is one of a text description

type, a phoneme description type and a formant frame description type, the text description type of the voice reproduction event data containing text information specifying words to be pronounced by the sound generator device as the human voice and associated prosodic symbols specifying vocal expression applied to pronunciation of the words, the phoneme description type of the voice reproduction event data containing phoneme information specifying phonemes of the human voice to be reproduced by the sound generator device and associated prosodic control information controlling vocal expressions of the phonemes, the formant frame description type of the voice reproduction event data containing formant control information specifying formants of the human voice at respective time frames.

16. A method of controlling a music apparatus having a data storage and a sound generator for reproducing a music sound and a voice sound, the method comprising the steps of:

storing a music data file containing a music part and a voice part in the data storage, the music part containing a sequence of music generation events effective to instruct generation of the music sound, the voice part containing voice reproduction sequence data composed of a combination of voice reproduction event data and duration data, the voice reproduction event data instructing reproduction of a sequence of voice events, the duration data specifying a timing of effecting a voice event in terms of a duration time measured from another voice event preceding to the voice event;

reading out the music data file from the data storage;

operating the sound generator based on the music part contained in the read music data file for generating the music sound representative of the sequence of the music events, and

operating the sound generator based on the voice part contained in the read music data file for generating the voice sound representative of the sequence of the vice events, thereby mixing and outputting the music sound and the voice sound.

17. A computer program for use in a music apparatus having a data storage and a sound generator, the computer program being executable in the music apparatus for performing a method of reproducing a music sound and a voice sound, wherein the method comprises the steps of:

storing a music data file containing a music part and a voice part in the data storage, the music part containing a sequence of music generation events effective to instruct generation of the music sound, the voice part containing voice reproduction sequence data composed of a combination of voice reproduction event data and duration data, the voice reproduction event data instructing reproduction of a sequence of voice events, the duration data specifying a timing of effecting a voice event in terms of a duration time measured from another voice event preceding to the voice event;

reading out the music data file from the data storage;

operating the sound generator based on the music part contained in the read music data file for generating the music sound representative of the sequence of the music events, and

operating the sound generator based on the voice part contained in the read music data file for generating the voice sound representative of the sequence of the vice events, thereby mixing and outputting the music sound and the voice sound.